

Introduction to the Postsecondary Education Transcript Studies (PETS)

Slide 1 of 61

Slide title: Introduction to the Postsecondary Education Transcript Studies (PETS)

Slide 2 of 61

This module will introduce users to the Postsecondary Education Transcript Studies (or PETS). It will describe the potential uses for the data and research questions appropriate for these data. It will also explain the data collections included in PETS and the associated data coding. Additionally, it will describe sources of data, data files, and sample variables. The module will also provide examples of the data sources (including publications and web tools like PowerStats), and data files. And finally, it will explain the data analysis appropriate for use with PETS data.

Throughout this module, underlined blue screen text indicates a link to additional resources.

Slide 3 of 61

Transcript data provide researchers with an important analytic resource, creating opportunities to study course-taking patterns, credit transfer, student momentum and attrition, and the connection among course and major choices, occupations, and wages.

Users who are interested in conducting analysis of PETS data are encouraged to follow the steps outlined in the following slides in order to ensure that the best approach is used.

Slide 4 of 61

Transcript collections are used for different purposes depending on the study. In secondary studies, for example, data from postsecondary transcripts can be used to examine pathways to a credential and the alignment between high school and college or postsecondary curricula. For secondary and postsecondary studies, it can be used to examine persistence, attainment, the timing of completion, alignment between courses taken and major, and if it is related to workforce outcomes, especially in Science, Technology, Engineering, and Mathematics (or STEM). Data can also be used to provide context associated with observations the student provides in the student interview.

Appropriate uses of PETS data depend on the purpose and scope of the study.

Slide 5 of 61

To date, transcript data have been collected from seven NCES surveys: the National Longitudinal Study of 1972 (or NLS-72), High School and Beyond 1980 (or HS&B:80), Baccalaureate and Beyond Longitudinal Study (or B&B:93/94), National Education Longitudinal Study of 1988 (or NELS:88), Beginning Postsecondary Students Longitudinal Study (or BPS:04/09), Baccalaureate and Beyond Study (or B&B:08/09) and Education Longitudinal Study of 2002 (or ELS:2002). Two additional data

Introduction to the Postsecondary Education Transcript Studies (PETS)

collections are the High School Longitudinal Study of 2009 (or HSL:09) and Beginning Postsecondary Students Longitudinal Study (or BPS:12/17).

Slide 6 of 61

The amount of detail in transcript collections varies depending on the study. Some studies, such as NELS:88, BPS:04/09, B&B:08, and ELS:2002 have detailed course, term, and institution records. Other transcript collections were used to verify interview data collected from the students in the study, (e.g., B&B:93/94). Researchers should pay special attention to how complete the transcript record is. In B&B:08, for example, transcripts were only collected for the final institution of attendance. And, in NELS:88, it is recommended to use the flag indicating that a complete transcript record was collected.

In the coming slides, this module will provide detailed information about transcript collections that occurred after 2010, including upcoming collections. While similar information is included in older studies, different coding and standardization methods were used in the more recent collections. Researchers should read the documentation for older studies to understand how they align or do not align with more recent collections.

Slide 7 of 61

Transcripts are typically collected directly from postsecondary institutions and are individually coded by trained keyer/coder staff. Response rates for these collections are typically higher than the response rates from student interviews. For example:

- BPS:04/09 collected transcripts from about 87% of institutions for about 87% of students and included records for approximately 17,000 students.
- B&B:08/09 collected transcripts from about 93% of institutions for about 94% of students and included records for approximately 17,400 students.
- ELS:2002 collected transcripts from about 92% of institutions for about 79% of students and included records for approximately 11,600 students.
- HSL:09 collected transcripts from about 55.5% of institutions for about 71% of students and included records for approximately 13,160 students.
- BPS:12/17 collected transcripts from about 68% of institutions for about 74% of students and included records for approximately 25,910 students.

Slide 8 of 61

Transcript information must be coded into useable data. In order to ensure that the data are efficiently and accurately coded, a five-step process is employed. First, definitions and coding rules must be developed.

For more information on definitions and coding, users should refer to The College Course Map (or CCM:2010) for more information. Second, extensive staff training must

Introduction to the Postsecondary Education Transcript Studies (PETS)

take place. Third, transcripts must be collected from institutions. Forth, the transcripts must be converted to electronic format. And fifth, quality control and spot checks must take place.

Slide 9 of 61

The College Course Map (or CCM) is a taxonomy system for coding postsecondary education courses in National Center for Education Statistics (or NCES) research studies. Earlier editions of the College Course Map were developed using faculty review groups in major disciplinary areas (e.g., mathematics, business, chemistry, allied health, and English).

These groups were provided descriptions from course catalogs and used NCES postsecondary transcript data to develop recommendations for coding and decision rules. The 2010 update to the CCM integrated selected courses from the CCM into the 2010 Classification of Instructional Programs (or CIP) taxonomy from NCES. CCM codes are represented by six digits in keeping with the CIP taxonomy: the first 2 digits indicate the most general category; the first 4 digits narrow the focus to a subcategory; and the complete 6-digit code provides the most specific definition of the subject.

Slide 10 of 61

To create a comprehensive course coder for PETS, content from the CCM was merged with the 2010 CIP and additional course codes were added during each study to accommodate technological developments since the CCM was developed. In this example, the description from 01.0601 mirrors the description of the CIP codes with additional instructions to the coders about what additional courses to include. To reflect a newer code not in previous versions of the CCM, 01.0698 was added to the course list at the end.

Slide 11 of 61

The College Course Map (or CCM) contains a set of 48 two-digit course categories. More detail about the CCM including a listing of all of the course codes can be found in the College Course Map Directory report.

Slide 12 of 61

A number of measures were taken to ensure quality in the PETS data collections. To evaluate consistency in the coding process, inter-rater reliability measures were reviewed daily. In a typical study, a percentage of transcripts are randomly selected for recoding and expert evaluation (typically 10%). In one joint collection of transcripts in two studies, this resulted in the expert evaluation of coding for 47,428 courses.

In addition, all un-codeable courses or courses listed as “other” were re-reviewed by expert coders from the field of study.

Introduction to the Postsecondary Education Transcript Studies (PETS)

Slide 13 of 61

Remedial courses can be identified using CCM codes. For example, the CCM 2-digit code for English is 23. The code 23.9987 indicates the course Remedial Speech, Basic Speech, Basic Oral Communication, and/or Listening Skills.

For Multidisciplinary Studies, the CCM 2-digit code is 30 so a course listed as 30.997 would indicate Basic Science and/or Remedial Science.

Slide 14 of 61

All remedial courses that are flagged in the derived variables of the dataset exclude audited courses or courses dropped before the withdrawal date, transfers, and advanced placement.

Slide 15 of 61

All course units were standardized to semester credits. For institutions that use clock hours, the clock hours were divided by 37.5. For semester, Advanced Placement (or AP) credit, College Level Examination Program (or CLEP) credit, study abroad credit, and International Baccalaureate (or IB) credit, the credit values were carried over unchanged. For quarter credit hours, quarter credits were multiplied by two-thirds.

Institutions with non-traditional credit systems were standardized based upon the instructions provided by the institution on the Transcript or in the student handbook for the institution. Sometimes additional investigation was required such as a document search of public information or a call to the registrar at the institution.

Slide 16 of 61

Course grades were standardized to a common 4-point system according to the chart on this slide. Further investigation was required for institutions with non-traditional grading systems.

Slide 17 of 61

In a typical study, there are six to nine datasets in PETS including "Institution," "Transcript," "Terms," "Courses," "Degrees/Major," "Student-Institution," or "Student Schools," "Transfer," and/or "Derived" (which may be combined with the Derived file with interview responses), and the "Weights." Each of these study components is explained in the upcoming slides.

Slide 18 of 61

Sample variables that may be found in PETS include measures such as:

- Postsecondary career credits earned,
- Credits earned by sector of institution,
- Credit attempted/earned by course subject,
- Total months enrolled,

Introduction to the Postsecondary Education Transcript Studies (PETS)

- Length of stop outs,
- First transfer variables,
- Elapsed time to award,
- Number of undergraduate courses,
- Credit earned by year, and
- Grade point average (GPA) by year.

Slide 19 of 61

To better understand how these variables are derived, let's examine each of the data sources and data files in the PETS collections. The "Institutions" data file includes information from three sources. The transcript (which include the header of the transcript, the documentation on the back of the transcript, or the materials sent with the transcript from the institution), data from the Integrated Postsecondary Education Data System (or IPEDS), and information from the Institution's course catalog.

Slide 20 of 61

This dataset includes all known institutions from which data are collected. Institutions that have Title IV eligibility, that is eligible for federal financial aid, have complete information carried over from the Integrated Postsecondary Education Data System or IPEDS, the course catalog, and transcripts. For other institutions including institutions outside of the United States, non-title IV institutions, un-codeable institutions (such as districts or system offices) or unknown institutions, much of this information is unavailable; although the name of the institution will be noted.

Slide 21 of 61

The variables included at this level are the sector, level of institution, control of institution, grading system, calendar system, clock vs. credit hours, and accreditation.

Slide 22 of 61

In order to assist researchers in selecting variables for analysis, this module will provide basic information about some variables in each dataset. Some variables are not meant to be used as analysis variables and require edits, computation, or need to be combined with other information/variables.

These variables are labeled with a red "X." Other variables have undergone this treatment and are represented here with a checkmark.

In this data file, the Institution ID, which is the IPEDS unit ID, can be used to match institution information to the student. Keep in mind, however, that some students attended multiple institutions. Additionally, while the state where the institution is located is included in the data file, the sampling design for many of the PETS studies is generally not representative at the state level.

Slide 23 of 61

The transcript data file is not included in all PETS studies. The Transcript data file draws most of its information from the transcript and the course catalog. Typically, data are taken from the transcript header and the information from each term. If a PETS study does not have a Transcript data file, this information can typically be found in the Student-Institutions file.

Slide 24 of 61

Cases included in the dataset are derived from transcript records.

Cases that are excluded from the dataset include transcripts not provided by the institution (i.e., course records may be in the courses dataset without a transcript record due to information learned on other transcripts).

Slide 25 of 61

Within the dataset for Transcript data, rows represent student-transcript records. Variables include the institution ID, high school graduation date, cumulative clock hours or credits, cumulative grade points earned, and the cumulative GPA.

Slide 26 of 61

Generally, very little information is available for analysis in this dataset without taking time to standardize the data. As shown on the screen, this dataset included unstandardized credit/clock hour and grade information, which may not be comparable across institutions.

Slide 27 of 61

The Terms dataset draws primarily from the transcript term records with supplemental information from the course catalog.

Slide 28 of 61

Transcripts or students with term records (most likely credit-hour institutions) are included in the dataset.

Cases that are excluded from the dataset include transcripts or students without term records (most likely from clock hour institutions).

Slide 29 of 61

For the Terms dataset, variables include, term start date, term end date, enrollment intensity, number of courses taken, credits attempted, credits earned, term GPA, and academic actions (e.g., Dean's list, or probation).

Slide 30 of 61

While the Terms dataset is good for providing supplemental information, the data reported in the dataset are not reliably reported by institutions. Any academic actions or classifications such as Dean's list, academic probation, and other classifications are reported but not by all institutions.

Some studies that report this information in the Terms dataset have computed variables that are reliable for use, including credits attempted, credits earned and GPA. Researchers should review the documentation to ensure they are selecting the correct variables.

Slide 31 of 61

Typically, the largest dataset in PETS is the Courses data file, which draws primarily on the transcript for information about courses. Course descriptions in the institution's course catalog are used to assist with accurately coding courses.

Slide 32 of 61

Rows in the courses dataset represent student-course records. Cases included in the Courses dataset include all known course records taken at each institution and/or transferred.

Cases that are excluded from the dataset include standardized test scores (e.g., AP or IB) although credits and bulk transfer credits will appear.

Bulk credit is when an institution listed transfer credit in one lump amount instead of itemizing the credit on a course level. For example, institution 1 reported that 30 hours of credit was accepted from institution 2 instead of listing the 10 courses accepted from institution 2. Bulk credit was reconciled when possible, meaning that the bulk credit was broken into the appropriate courses when possible. For example, if 6 hours were listed as bulk credit on transcript #1 and 2 3-hour courses were listed on transcript #2, the credit was associated with the 2 3-hour courses. This provided more detail for analysis. In the event that bulk credit could not be reconciled, the credit was left intact.

Slide 33 of 61

Courses data may be duplicated (or listed multiple times) due to a student transferring institutions because course records can be included on more than one transcript. The course variables include course name, course number, PETS/CCM code for the course, earned credit/clock hours, standardized credit/clock hours, earned grade, standardized grade, and/or course attributes and flags. Examples of course attributes include lab course, honor course, remedial course, or course format.

Note that newer PETS studies include blank course records to indicate that records are missing due to known transcripts that were not submitted by an institution.

Slide 34 of 61

Most variables in the courses dataset are not standardized and are directly pulled from the transcript. Therefore, they should not be used in analysis unless they are standardized across institutions. Examples include credits, clock hours, earned credits, or grade. In order to allow for the derivation of variables in other datasets, the Courses dataset includes some standardized variables such as standardized credits received for a course standardized grade, quality points, or potential credits earned. There are some course variables or course attributes that include two versions. One is a raw variable copied directly off the transcript and the second is a computed variable.

Slide 35 of 61

For example, the courses dataset includes a transfer course credit flag, which is not reliably reported on all institution transcripts. A second version of the variable, which is derived by matching course records across transcript, is included in the courses dataset.

Other variables not reliably reported on transcripts include attributes such as remedial, research, seminar, study abroad courses, and others. If it was possible to derive these measures other ways, they are included in the Derived dataset. The most important variable on the Courses data file is the CCM course code assigned to the course by the course coder.

Slide 36 of 61

The Degrees/Majors data file includes information from the transcript and course catalog. If present, major information is acquired from term records, and degree and major information is recorded from a notation typically at the beginning or end of a transcript record.

Slide 37 of 61

For degrees/major data in the dataset, rows represent student-degree or student-major records.

Students with degrees or certificates (duplicate records possible) are included in these data, while students without a degree/certificate or major are excluded.

In other words, if a student earned two degrees, a certificate and a degree, or if a major was changed, a student can have multiple rows in the file. For example, if a student earned a certificate in year 2, and a bachelor's degree in year 5, and it is unclear from the transcript what the student's major was, the Degrees/Majors dataset would have two records. For another student, if the institution has a lot of major information on the transcript and the student changed her/his major once before getting a degree, s/he would have multiple records, one for each major and one for the degree. It is important to note that degree/major information is not reliably reported on institution transcripts unless an award is conferred.

Slide 38 of 61

Variables included from these data are degree program, date of degree, units required for degree, major, second major, minor, second minor, concentrations, notations and honors.

Slide 39 of 61

The student institution data file includes all known institutions that a student attends or attended. Data for this file are primarily from institution transcripts. When an institution does not respond to NCES' request for transcripts, the information is gathered from other sources, including transcripts from other institutions.

For example, as shown on the screen, the University of A transcript provided data for the Community College of B and C State University in the Student Institutions dataset. Older PETS files call this the "Student Schools" dataset.

Slide 40 of 61

For student institution data, the rows represent student-institution records and there may be multiple records for each student. The dataset included student/institutions pairs. Unknown and missing institutions are included in the dataset but with missing data.

Slide 41 of 61

For student institution data, multiple records for each student may be included. It should be noted that if a student transfers out of an institution and eventually returns to the target institution, they will have one record. In this example, when a student transfers from institution A to B, then from B to C, and then back to A, there will be 3 student/institution records in the dataset even though there was a break in attendance at institution A.

Slide 42 of 61

The resulting record will be similar to what is seen on this table. The dataset includes the order of attendance.

In this example, the order of attendance is institution A, B, then C based on the start dates and B, C, then A based on the end dates. Extra effort was made to ensure that the dates of attendance were recorded. If the dates were missing from a transcript, then dates were identified using the Course Catalog or an institution website.

Slide 43 of 61

Student-Institution variables in the dataset include ratio of credits earned to credits attempted, number of terms enrolled, the number of terms enrolled full-time, transfer credit attempted, and transfer credit accepted.

Slide 44 of 61

The majority of the Student-Institutions dataset are available for use without editing the variable such as credits earned, credits earned by subject, grade point average, and grade point average by subject. Caution should be used, however, when using the transfer variables. Many of the transfer credit/GPA variables only included data from transcripts where credits are itemized. Many institutions will note that credits were transferred, but do not indicate which credits or courses were recognized. NCES classifies these as bulk credits transferred and bulk credits are not included in many of the transfer variables.

As a result, a second set of transfer variables are added to the dataset. The only measure that can be accurately derived for transferring students is the number of credits transferred in a student's first transfer. This is because institutions that indicate bulk credits on the transcript may not itemize where the credits came from. Most other variables on the dataset are acceptable to use.

Slide 45 of 61

For some studies, a special dataset is available so researchers can examine student transfer. The Transfer dataset uses information for all the students' transcripts collectively.

For institutions that itemize credits and not all the previous institutions of attendance, all this information is included in the transfer dataset. For institutions that only list bulk credit or do not indicate where credits come from, all other transcripts are examined to deduce a student's transfer history.

Slide 46 of 61

For Transfer data, the rows represent student-origin, institution-destination, institution records. Transfer data include opportunities for credit transfer but they do not include transfer opportunities with missing and/or end dates. When using this dataset, researchers should use the Student-Institutions dataset to track which records were excluded due to missing beginning or end dates.

Newer studies will include reserve codes in the Transfer dataset. This dataset is constructed using the definition of a credit transfer opportunity, which will be explained on the next slide.

Slide 47 of 61

A potential transfer credit opportunity is an opportunity for credits to move between institutions as a result of a student attending multiple institutions. Potential transfer opportunities are identified using the beginning and end dates of attendance at each institution. The more institutions a student attends, the more institution-to-institution relationships can be established, thus leading to more potential transfer opportunities.

Additional opportunities can be documented when students have overlapping dates of attendance because credits could be transferred in either direction.

Slide 48 of 61

Typically, students do not take advantage of all potential transfer credit opportunities. The illustration on the screen shows credits from Institution B and institution C transferred to Institution A but credits did not transfer from A to the others or between B and C.

For example, a student who transfers three times with no overlapping dates of attendance, as shown in the illustration on the screen, has three student transfers: transfer one (or T1) shows the move from institution A to institution B with a potential transfer opportunity represented by dotted lines; transfer two (or T2) shows the move from institution B to institution C with two potential transfer opportunities from institution A to C and C to D; and transfer three (or T3) shows the student's move from institution C back to institution A with two additional potential transfer opportunities back from C to A and from B to A. These transfers constitute six potential transfer opportunities, which increase the more institutions the student attended.

Slide 49 of 61

Each row in the Transfer dataset represents a Potential Transfer Opportunity, regardless of whether credits moved or not.

Slide 50 of 61

The variables included in the Transfer data are beginning and end dates, first transfer flag, total credits taken at original institution, total credits accepted at destination institution, percentage of credits transferred, sector relationship, accreditation relationship, and degree program change flag.

Slide 51 of 61

Most variables in the Transfer dataset are available for use, especially variables that document the movement between institutions such as the

- Accreditation relationship (e.g., regionally accredited to nationally accredited)
- Control relationship (e.g., public to private nonprofit, public to private for-profit, etc.), and
- Level relationship (e.g., 2-year to 4-year, 4-year to 4-year, etc.).

An "Incidental Transfer" flag is also included in the dataset. For students who attended more than one institution, "incidental transfer" refers to an enrollment spell of less than 4 months at a second (or subsequent) institution of attendance with a return to the origin institution.

Any variable related to a student's program of study or "major" is generally not useable. A student's major is not reliably reported on transcripts. If, however, a student earns a degree or certificate, major is reported. Because there are multiple variables with similar names on the dataset, researchers should take extra care to read the documentation.

Slide 52 of 61

The Derived dataset compiles and combines all information into a data file that analysts can use for research. It includes data from multiple sources including student interviews, administrative data, and transcripts. Some older studies have a derived data file that is separate from the derived data file based on the student interview. In such studies, these files can be combined. For the Derived data file, each row represents a student.

Slide 53 of 61

For Derived data, variables include postsecondary career credits earned, credits earned by sector of institution, credit attempted/ earned by course subject, total months enrolled, and length of stop-outs.

In addition, the 1st transfer variables (for select studies) include origin credits, credits transferred, percent transferred, selectivity relationship, accreditation relationship, and credits transferred by sector. Additional variables include: elapsed time to award, number of undergraduate courses, credit earned by year, and GPA by year.

Slide 54 of 61

Most variables in the Derived dataset are recommended for use. Because there is no imputation in transfer variables, researchers should pay close attention to missing cases. In this case, the Major variable has high levels of missing-ness, so researchers should supplement major information or exclude the variable in models employing listwise deletion.

All summed or computed variables based on counts, credits attempted, credits earned, and ratios are also reliable for use. The credit transfer indicator is also acceptable to use, but researchers should keep in mind that it only includes the first student transfer.

Slide 55 of 61

As shown in the previous slides, transcript data provide researchers with an important analytic resource. PETS data are appropriate for addressing research questions about course-taking patterns, credit transfer, student momentum and attrition, and the connection among course and major choices, occupations, and wages.

It is important to keep in mind, that analysis of PETS data requires access to specific micro-data files through a Restricted-use license. Use of these files also requires researchers to understand and apply weights and design variables to ensure proper analysis. For more information about analysis of NCES complex datasets, review the NCES Computer-Based Training modules "Analyzing NCES Complex Survey Data and Statistical Analysis of NCES Datasets Employing a Complex Sample Design". Another way to access PETS data is through the data tool PowerStats (public use or restricted use datasets without having to obtain a restricted use license).

Introduction to the Postsecondary Education Transcript Studies (PETS)

Users who are interested in conducting analysis of PETS data are encouraged to follow the steps outlined in the following slides in order to ensure that the best approach is used.

Slide 56 of 61

A critical decision for using PETS data includes selecting the most appropriate dataset. For a given research question, the derived dataset through PowerStats should be used first.

Slide 57 of 61

If the derived dataset cannot answer the research question, NCES recommends using Student Schools and/or the Transfer datasets.

Slide 58 of 61

If the derived, student schools, weights, and transfer datasets cannot answer your research question, use the Institution, Transcript, Terms, Courses, and/or Degrees datasets.

Note, though, that these data require that files are merged, and unstandardized/non-standardized variables must be converted to a useable format.

Slide 59 of 61

Variables from the Derived files are available on PowerStats for use by researchers. PowerStats does much of the work for the researcher as it enables users to access results based on unit record NCES data without utilizing a statistical analysis program. In other words, PowerStats is a table and statistics generator that uses Restricted-use data without the need for a license.

Slide 60 of 61

This module introduced users to the Postsecondary Education Transcript Studies (or PETS).

It described the potential uses for the data and research questions appropriate for these data. It also explained the data collections included in PETS and explained the associated data coding. Additionally, it described sources of data, data files, and sample variables. The module also provided examples of the data sources/files. And finally, it explained the data analysis appropriate for use with PETS data.

Slide 61 of 61

Important resources that have been provided throughout this module are summarized on this slide for your reference.

You may now click the “Exit” button to return to the landing page.